**ISO/IEC 14496-15:20xx(E)**

ISO/IEC jTC 1/SC 29/WG 03

Date: 2023-02

**Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format — Amendment 2: Picture-in-picture support and other extensions**

Draft text of DAM stage

**Warning for WDs and CDs**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

© ISO 20XX

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO’s member body in the country of the requester.

ISO copyright office

CP 401 • Ch. de Blandonnet 8

CH-1214 Vernier, Geneva

Phone: +41 22 749 01 11

Email: copyright@iso.org

Website: www.iso.org

Published in Switzerland

Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format — Amendment 2: Picture-in-picture support and other extensions

*Clause 2*

Replace the list of normative references with the following: [Ed. (YK): Note that in addition to adding 23008-12, I also removed the year from the references for 14496-12, 23090-3, and 23094-1. Pointing this out just in case removing the year for those is not good. (MH): Dated references are required when specific clause numbers, Tables, Figures or Equations are referenced (e.g., "as specified in ISO/IEC 14496-12, subclause X.Y"). I have had the practice to include dates in Clause 2 whenever dated references for that standard appear in the text, but I am not sure if that is necessary.]

IETF RFC 4648, *The Base16, Base32, and Base64 Data Encodings*

IETF RFC 6381, *MIME Codecs and Profiles*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 23008-2, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

ISO/IEC 23008-12, *Information technology — MPEG systems technologies — Part 12: Image file format*

ISO/IEC 23090-3, *Information technology —* *Coded representation of immersive media — Part 3: Versatile video coding*

ISO/IEC 23094-1, *Information technology —* *General video coding — Part 1: Essential video coding*

*Subclause 3.1.6*

Replace the definition of AVC NAL unit with the following:

**AVC NAL unit**

AVC VCL NAL unit or a non-VCL NAL unit associated with an AVC VCL NAL unit as specified in ISO/IEC 14496-10

*Subclause 4.2.3.3*

*In subclause 4.3.3.3, replace*

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in the appropriate specification (e.g. ISO/IEC 14496-10) and includes both the one byte NAL header and the variable length encapsulated byte stream payload.

*with*

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in the appropriate codec specification (e.g. ISO/IEC 14496-10) and includes both the NAL unit header and the variable length NAL unit payload.

*New subclauses 4.16 to 4.18*

*Add subclauses 4.16 to 4.18 as follows:*

* 1. **Storage of SEI manifest and SEI prefix indication SEI messages**

When one or more SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message need to be stored, they should be stored in the decoder configuration information as follows:

* For AVC, SVC, and MVC, these SEI NAL units should be stored as instances of sequenceParameterSetNALUnit in AVCDecoderConfigurationRecord, SVCDecoderConfigurationRecord, and MVCDecoderConfigurationRecord, respectively.
* For HEVC and L-HEVC, these SEI NAL units should be stored as instances of nalUnit with NAL\_unit\_type indicating a prefix SEI NAL unit in HEVCDecoderConfigurationRecord and LHEVCDecoderConfigurationRecord, respectively.
* For VVC, these SEI NAL units should be stored as instances of nal\_unit with NAL\_unit\_type indicating a prefix SEI NAL unit in VvcDecoderConfigurationRecord.

When a parameter set elementary stream is in use for AVC, SVC, MVC, or MVD, SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message should also be placed in the parameter set elementary stream.

* 1. **Supplementary track reference**

[Ed. (YK): The features specified in this subclause are generic and should be moved to 14496-12 in the future.]

Multimedia applications may offer a functionality, sometimes referred to as picture-in-picture service, of displaying a video with a smaller spatial resolution over a video with a bigger spatial resolution. In this context, the videos with the smaller and bigger spatial resolutions referred to as the supplementary video and the main video, respectively.

ISO/IEC 14496-12 supports such a functionality when an explicit association between a pair of main video and supplementary video is not necessary, e.g., when the file contains only two video tracks, by providing the size, position and layer of the main video and supplementary video through the header of their respective tracks. Regardless of whether the association between a pair of main video and supplementary video is explicitly signalled, for each pair of main video and supplementary video, the window in the main video for overlaying the supplementary video, which is smaller in size than the main video, is indicated by the values of the matrix fields of the TrackHeaderBoxes of the supplementary video track and the main video track, and the value of the layer field of the TrackHeaderBox of the supplementary video track shall be less than that of the main video track, to layer the supplementary video in front of the main video.

In some cases, a file writer may want or need to explicitly signal the association between a pair of main video and supplementary video. For example, when there are more than three video tracks in the file while only two of the video tracks are intended to be used together to provide a combined experience, an explicit signalling of the association is needed. For this purpose, the 'supm' track reference is defined and indicates that the video in the referencing track or in any track in the alternate group to which the referencing track belongs, if any, can be used as a supplementary video, and that the corresponding main video can be the video contained at least in a referenced track or any track in the alternate group to which the referenced track belongs, if any.

A track with a track reference of type 'supm' shall have a handler 'vide' (specified in ISO/IEC 14496-12) or 'pict' (specified in ISO/IEC 23008-12).

For each pair of supplementary video and main video, the following applies:

* The track containing the supplementary video is also referred to as the supplementary video track.
* The main video track is the track that is the referenced track of the 'supm' track reference or any track in the alternate group to which the referenced track belongs, if any.
* When the video codec used for a main video is HEVC or L-HEVC, the main video may be contained either in a single track or in multiple tracks consisting of an HEVC or L-HEVC tile base track and a number of HEVC or L-HEVC tile tracks containing 'tbas' track references that refer to the tile base track. In the former case, the main video track is that single track. In the latter case, the main video track is the tile base track.
* When the video codec used for the main video is VVC, the main video may be contained either in a single track or in multiple tracks consisting of a VVC merge base track and a number of VVC subpicture tracks referenced by the VVC merge base track through the 'subp' track reference. In the former case, the main video track is that single track. In the latter case, the main video track is the VVC merge base track.
  1. **Picture region replacement sample group**

### Definition

When PicRegionReplacementEntry is present in a supplementary video track, it indicates that the NAL units representing the target supplementary video region in the main video can be replaced with the corresponding NAL units of the supplementary video. In this case, the following are required: 1) The same video codec is used for coding of the supplementary video and the main video; 2) The to-be-replaced region shall be coded with NAL units containing the data and only data for the corresponding region, and no data outside the region references data within the region; and 3) The replacement of these NAL units by the NAL units of the supplementary video track (sample by sample) produces a bitstream compliant to the one indicated by the sample entry. The absence of this sample group indicates that it is unknown whether such replacement is possible.

When this sample group is present, the player may choose to replace the NAL units representing the target supplementary region in the main video with the corresponding NAL units of the supplementary video before sending to the video decoder for decoding. In this case, for a particular picture in the main video, the corresponding NAL units of the supplementary video are all the NAL units in the decoding-time-synchronized sample in the supplementary video track.

### Syntax

class PicRegionReplacementEntry() extends VisualSampleGroupEntry ('pprr') {  
 bit(5) reserved = 0;  
 unsigned int(3) region\_id\_type;  
 unsigned int(8) num\_region\_ids\_minus1;  
 for(i=0; i<=num\_region\_ids\_minus1; i++)  
 unsigned int(16) region\_id[i];  
}

### Semantics

region\_id\_type indicates the type for the value taken by the region\_id. If the video codec used for the main video track is VVC (i.e., the sample entry type is 'vvc1' or 'vvi1'), in which case the video codec used for the supplementary video track is also VVC, region\_id\_type equal to 0 specifies that the region IDs are VVC subpicture IDs. Otherwise, the value of region\_id\_type equal to 0 is reserved. When region\_id\_type is equal to 1, the region IDs are the groupID values in the NAL unit map sample group for the NAL units that may be replaced by the NAL units of the supplementary video track. region\_id\_type values greater than 1 are reserved.

num\_region\_ids\_minus1 plus 1 specifies the number of the following region\_id[i] fields.

NOTE The number of regions can be restricted by a specification that uses the picture region replacement sample group through referring to this document.

region\_id[i] specifies the i-th ID for the NAL units representing the target picture region.

The following constraints apply:

* When region\_id\_type is equal to 1, the main video track shall have a 'nalm' sample group with grouping\_type\_parameter equal to 'pprr' indicating the NAL units in the main video that may be replaced by the NAL units in the supplementary video track with the same groupID values.
* When region\_id\_type is equal to 1 and num\_region\_ids\_minus1 is equal to 0, a 'nalm' sample group shall not be present in the supplementary video track and all the NAL units of the supplementary video track are inferred to have groupID equal to region\_id[0].
* When region\_id\_type is equal to 1 and num\_region\_ids\_minus1 is greater than 0, a 'nalm' sample group with grouping\_type\_parameter equal to 'pprr' shall be present in the supplementary video track and provide a mapping of groupID values to NAL units.

*Subclause 5.4.3.2.3*

*In subclause 5.4.3.2.3, replace*

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in ISO/IEC 14496-10 and includes both the one byte NAL header and the variable length encapsulated byte stream payload.

*with*

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in ISO/IEC 14496-10 and includes both the NAL unit header and the variable length NAL unit payload.

*Subclause 6.4.2.2*

*In subclause 6.4.2.2, replace*

SequenceParameterSetLength indicates the length in bytes of the SPS or subset SPS NAL unit.

SequenceParameterSetNALUnit contains a SPS or subset SPS NAL unit. SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Subset SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Any SPS shall occur before all the subset SPSs, if any.

*With (just changing the starting ‘S’ with ‘s’ in both sentences, no other changes)*

sequenceParameterSetLength indicates the length in bytes of the SPS or subset SPS NAL unit.

sequenceParameterSetNALUnit contains a SPS or subset SPS NAL unit. SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Subset SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Any SPS shall occur before all the subset SPSs, if any.

*Subclause* 7.6.3.1.1.3

Replace the following:

Each component of the intrinsic matrix is obtained from the variables specified in Table 6 as the variable x computed as follows.

– If 0 < e < 63, x = 2e-31 \* (1 + n  2v), with v = max(0, e + p - 31) [Eq. F-1]

– If e is equal to 0, x = 2-(30+v) \* n, with v = max(0, p - 30) [Eq. F-2]

with the following:

Each component of the intrinsic matrix is obtained from the variables specified in Table 6 as the variable x computed as follows.

– If 0 < e < 63, x = 2e-31 \* (1 + n  2v), with v = max(0, e + p - 31)

– If e is equal to 0, x = 2-(30+v) \* n, with v = max(0, p - 30)

*Subclause* 7.6.3.1.2.3

Replace the following:

Each component of the rotation matrix and the translation vector is obtained from the variables specified in Table 7 as the variable x computed as follows.

– If 0 < e < 63, x = 2e-31 \* (1 + n  2v), with v = max(0, e + p - 31) [Eq. F-3]

– If e is equal to 0, x = 2-(30+v) \* n, with v = max(0, p - 30) [Eq. F-4]

with the following:

Each component of the rotation matrix and the translation vector is obtained from the variables specified in Table 7 as the variable x computed as follows.

– If 0 < e < 63, x = 2e-31 \* (1 + n  2v), with v = max(0, e + p - 31)

– If e is equal to 0, x = 2-(30+v) \* n, with v = max(0, p - 30)

*Subclause* 7.7.6.1

Replace the following:

NOTE: A Multiview Scene Information SEI message, as specified in MVC H.12.1.5, can indicate the maximum disparity between any adjacent views in the bitstream. Thus, the Multiview Scene Info Box represents similar information as carried in the Multiview Scene Information SEI message but is limited to a certain set of views rather than concerns all the views in the bitstream.

with the following:

NOTE: A multiview scene information SEI message, as specified in ISO/IEC 14496-10:2020, subclauses H.13.1.4 and H.13.2.4, can indicate the maximum disparity between any adjacent views in the bitstream. Thus, the Multiview Scene Info Box represents similar information as carried in the multiview scene information SEI message but is limited to a certain set of views rather than concerns all the views in the bitstream.

*Subclause 8.3.2.1.1*

*In subclause 8.3.2.1.1, replace*

* The level indication general\_level\_idc shall indicate a level of capability equal to or greater than the highest level indicated for the highest tier in all the parameter sets.

*with*

* The level indication general\_level\_idc shall indicate a level of capability greater than or equal to the highest level in all the parameter sets.

*Subclause 8.3.2.1.3*

*In subclause 8.3.2.1.3, replace*

array\_completeness when equal to 1 indicates that all NAL units of the given type are in the following array and none are in the stream; when equal to 0 indicates that additional NAL units of the indicated type may be in the stream; the default and permitted values are constrained by the sample entry name.

*with*

array\_completeness when equal to 1 indicates that all NAL units of the given type are in the following array and none are in the stream; when equal to 0 indicates that additional NAL units of the indicated type may be in the stream; the permitted values are constrained by the sample entry name.

*In subclause 8.3.2.1.3, remove the following from the semantics of nalUnit:*

When one or more SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message are available, they should be stored as instances of nalUnit.

*Subclause 8.4.1.1.1*

*In subclause 8.4.1.1.1, replace*

When the sample entry name is 'hvc1', the default and mandatory value of array\_completeness is 1 for arrays of all types of parameter sets, and 0 for all other arrays. When the sample entry name is 'hev1', the default value of array\_completeness is 0 for all arrays.

*with*

When the sample entry name is 'hvc1', the value of array\_completeness shall be equal to 1 for the arrays of all types of parameter sets.

*Subclause 9.5.3.1.1*

*In subclause 9.5.3.1.1, replace*

When the sample entry name is 'lhv1', the default and mandatory value of array\_completeness is 1 for arrays of all types of parameter sets, and 0 for all other arrays. When the sample entry name is 'lhe1', the default value of array\_completeness is 0 for all arrays.

*with*

When the sample entry name is 'lhv1', the value of array\_completeness shall be equal to 1 for the arrays of all types of parameter sets.

*Subclause 9.5.5*

Remove NOTE 2 and relabel NOTE 1 as NOTE.

*Subclause A.7.1*

*In subclause A.7.1, replace*

An extractor contains one or more constructors. The following constructors are specified:

1. A sample constructor extracts, by reference, NAL unit data from a sample of another track.
2. An in-line constructor includes NAL unit data.

*with*

An extractor contains one or more constructors. The following constructors are specified:

1. A sample constructor extracts, by reference, NAL unit data from a sample of another track.
2. An in-line constructor includes NAL unit data.
3. A sample constructor SampleConstructorFromTrackGroup extracts, by reference, NAL unit data according to a copy mode from a sample of another track or a track selected from a track group.
4. A reference constructor allows including a default constructor by explicit reference in an extractor.
5. A default reference constructor allows including a default constructor by implicit reference in an extractor.
6. A NALU start in-line constructor includes beginning of a NAL unit data without NALUnitLength field.

*Subclause A.7.6.2*

*Replace the content of subclause A.7.6.2 with the following (adding the semantics of* sample\_offset *and* nalu\_idx\_field\_size*, no other changes):*

track\_ref\_index specifies the index of the track reference of type 'scal' to use to find the track\_ID or the track\_group\_id from which to extract data. When the track\_ref\_index resolves to a track\_group\_id, it is up to the parser or player to select the most appropriate track in the corresponding track group depending on the track\_group\_type. A default behaviour is to select the first track in the file having the specified track\_group\_id.

sample\_offset: as specified in subclause A.7.4.1.2.

copy\_mode: specifies the copy operation to be performed when resolving the extractor:

**—** When set to 0, it means a sample copy, i.e. a copy of bytes from the first byte of the sample until the end of the sample, inclusive.

**—** When set to 1, it means a NAL unit copy, i.e. a copy from the first byte of the i-th NAL unit to the last byte of this same NAL unit, where i corresponds to the nalu\_idx field.

**—** When set to 2, it means a NAL unit payload copy; i.e. a copy from the first byte immediately following the NAL unit header in the i-th NAL unit payload to the last byte of this same NAL unit, where i corresponds to the nalu\_idx field.

**—** copy\_mode 3 is reserved for future use.

NOTE copy\_mode 2 is useful when some header rewriting is performed. In such case, only NALU payload is extracted and combined with rewritten NALU header, e.g. when some NALUs from different IRAP and non-IRAP pictures are merged in one single picture, there can be a need to rewrite nal\_unit\_type in NALU headers.

nalu\_idx\_field\_size: specifies the length of the nalu\_idx field. If the value is 0, the length of the nalu\_idx field is 8 bits, otherwise the length of the nalu\_idx field is 16 bits.

nalu\_idx: 1-based index of the NAL unit from where to extract. Value 0 is reserved. NAL-unit-like structures and NAL units that are present in the sample and have nal\_unit\_type value in the range of 48 to 63, inclusive, shall not be accounted for. NAL units included or referenced by an Aggregator shall be accounted for.